Articulatory-Acoustic Relationships in Fricative Consonants

C.H. Shadle
VSSP Technical Report 89/TR3

16 August, 1989

Department of Electronics and Computer Science

University of Southampton

Southampton S09 5NH

Abstract

The work described in this Report is concerned with improving fricative models by investigating the acoustic mechanisms involved in their production more closely. Previous studies reviewed fall into four categories: general aeroacoustics, mechanical models of speech, analysis of speech, and theoretical models, including synthesis. The work reported here combines these various approaches. Three levels of experiments with mechanical models of increasing realism are described. As a result, two source mechanisms are identified: the *obstacle* source, in which sound is generated at an abrupt obstacle to the airflow, such as the teeth, downstream of the constriction, corresponding to /s, ʃ/; and the *wall* source, in which sound is generated by striking a wall at an oblique angle to the flow, corresponding to /ç, x/. Source spectra differ, and are given in each case. Comparison with speech data from which the models were derived supports and extends the conclusions.

Contents

1 Introduction 1
2 Previous Work 2
2.1 General aspects of unstable and turbulent jets 2
2.2 Mechanical model experiments 3
2.3 Theoretical models 4
2.4 Speech analysis 5
3 Level I Models 5
3.1 Method 5
3.2 Results 7
3.3 Comparison with speech 7
4 Level II Models 10
4.1 Method 10
4.2 Results 11
5 Level III Models 11
5.1 Method 11
5.2 Results 14
5.2 a Source location 14
5.2 b Source characteristic 16
5.3 Comparison with speech 18
6 Conclusion 18
7 Acknowledgements 20
8 References 20

1 Introduction

A fricative consonant is produced when the vocal tract is constricted somewhere along its length enough to produce a noisy sound when air is forced through the constriction. As with vowels, the location of the constriction affects the timbre of the resulting sound. In addition, the vocal folds may vibrate simultaneously, generating a periodic sound at the glottis and modulating the airflow through the constriction. Thus we have, in English, the voiced-voiceless pairs of fricatives, arranged by constriction location, from most anterior to most posterior: /v, f; ð, θ; z, s; ʒ, ʃ/, and the voiceless glottal fricative, /h/ (these are, respectively, the italicised portions of *v*ine, *f*ine; *th*ine, *th*in; *z*ip, *s*ip; a*z*ure, a*ss*ure and *h*eat). Figure 1 illustrates a range of constriction locations by showing schematised mid-sagittal cross-sections of the vocal tract for the three fricatives /f, ʃ, x/ (as in *f*in, *sh*in, lo*ch*); in each case, an arrow points to the region of greatest constriction.

Fig. 1: Diagram of the mid-sagittal cross-section of the vocal tract during the production of the fricatives /f, ʃ, x/.

Fricatives are primarily classified by presence of voicing and location of the constriction. Finer classifications have been suggested, dividing fricatives into sibilants and non-sibilants, or stridents and non-stridents. The lack of agreement regarding which feature is more useful, or indeed which fricatives ought to be in which group, reflects our poor understanding of the acoustic mechanism of fricatives.

This poor understanding exists for three reasons. First, there is not a complete theoretical formulation of sound generation due to turbulence. Second, there is no mechanical vibration that is obviously correlated with the speech signal, as there is in the case of vowels; thus the primary sound generation process is more difficult to measure and to model physically. Third, the output signal – that is, the speech – is intrinsically noisy, and thus must be described statistically rather than analytically. Two signals might thus be produced by the same configuration under conditions considered identical in all practical terms, and yet appear quite dissimilar in the time domain. The process of comparing, coding, or analyzing speech signals is more complex; the models of the signals and the systems which produce them must also be more complex.

Past work on fricatives has been conducted using one or more of the following three approaches:

- mechanical model experiments;
- theoretical models, sometimes including synthesis and perception experiments;

- speech analysis, often including extensive articulatory measurements.

We will take these up in turn, after reviewing relevant aeroacoustics studies.

2 Previous Work

2.1 General aspects of unstable and turbulent jets

Techniques such as flow visualisation have established that as air exits from a constriction it forms a jet, which gradually mixes with the surrounding air. The Reynolds number (*Re*) characterises the degree of turbulence generated as this mixing takes place. It is defined by:

$$Re = \frac{vd}{\nu}$$

where v is a representative flow velocity, usually taken to be that in the centre of the constriction exit, d is a representative dimension, usually the constriction diameter, and ν is the kinematic viscosity of the fluid, which for air is 0.15 cm^2/s. As *Re* increases, an initially laminar flow will pass through an unstable region and finally become fully turbulent. Turbulent flow is distinguished by irregular, high-frequency fluctuations in velocity and pressure at a given point in space (Schlichting, 1979). The critical Reynolds numbers, Re_{crit}, separating these regions vary according to the geometry and degree of prior turbulence of the fluid. For a jet issuing from a circular hole, the unstable region would typically occur for $160 < Re < 1200$ (Goldstein, 1976).

The dimensions of a fully turbulent jet in the subsonic range depend only on the constriction diameter and shape; thus, visually, all jets can be scaled to look the same. Theoretical work by Lighthill (1954) and others established that the sound generated by jets scales as well, that is, that the spectral characteristics of the sound generated by a jet depend only on the jet velocity and diameter.

Sound is generated by the random pressure fluctuations of the turbulent fluid. A good summary of the theoretical and empirical efforts to describe this sound generation process may be found in Goldstein (1976). For our purposes, the essential facts are as follows. For a jet emerging from a constriction of diameter d at $Re > Re_{crit}$, three regions, the *mixing*, *transition* and *fully developed* regions, can be defined, as shown in Figure 2. From both theory and experiment, it appears that nearly all of the sound power is generated in the mixing and transition regions, possibly with most of it coming from the mixing region (Goldstein, 1976). If half of the sound power is assumed to be generated in the mixing region, the total power, P, generated by the jet is proportional to v^8, which agrees with Lighthill's prediction (1952). The total sound power spectrum has a broad peak at about Sv/d Hz, where v is the flow velocity in the centre of the jet as it exits the constriction, d is the jet diameter, and S, the Strouhal number, defined by this equation, is equal to 0.15. (The frequency of the spectral peak depends on the type of spectrum chosen. The Strouhal number at the peak is $S = 0.15$ when the noise spectral density, an equal-bandwidth representation, is plotted; $S = 1.0$ for the third-octave spectrum.) The sound pressure measured at a particular point in the far field will have a similar spectrum, with a peak frequency dependent on the angle at which the measurement is made. Measurements within the jet itself show that the high-frequency sound originates closer to the nozzle than does the low-frequency sound (Fletcher and Thwaites, 1983).

Lighthill described three types of sound sources that are present to varying degrees in the sound produced by turbulent flow: a monopole source (which is equivalent to a sphere pulsing in and out), a dipole source (two spheres pulsing in opposite phase), and a quadrupole source. A monopole source obeys a v^4 power law, meaning that the total sound power generated by a flow monopole increases as the fourth power of the

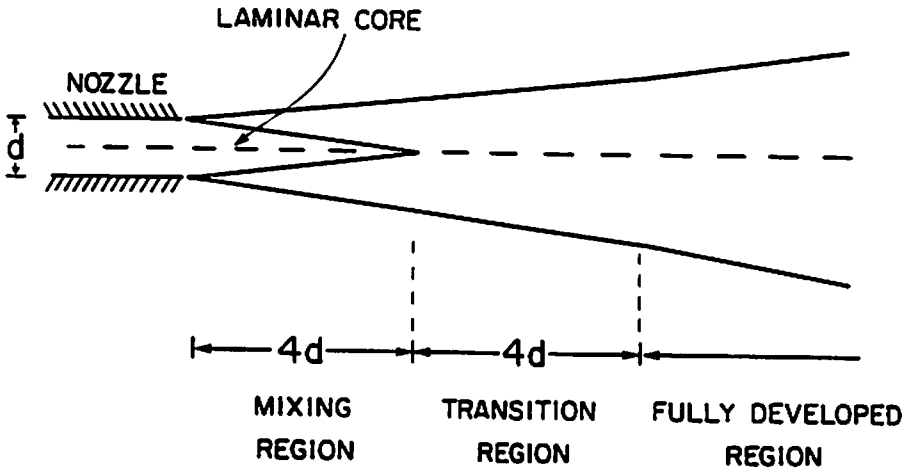

Fig. 2: Diagram of the mixing, transition and fully developed regions of a fully turbulent jet, after Goldstein (1976).

flow velocity v; the dipole source obeys a v^6 power law; the quadrupole source, v^8. An equivalent statement is that the efficiency of conversion of the kinetic energy of the flow into sound is proportional to $M = v/c$ for the monopole, M^3 for the dipole, and M^5 for the quadrupole source (Morse and Ingard, 1968). For subsonic velocities ($M < 1.0$) the quadrupole is thus the least efficient source, but its relative contribution to the total sound generated should become progressively more important as M increases. Dipole sources occur along rigid boundaries, which exert an alternating force on the fluid; quadrupole sources exist in free jets. Thus if a jet impinges on an obstacle such as a flat plate, the sound generation can be modelled by a combination of quadrupole and dipole sources. It is not well understood how the energy divides between quadrupole and dipole sources in such a case (Goldstein, 1976).

Two studies have combined theory and experiment to focus on the sound generated by a spoiler in a duct. Heller and Widnall (1970) showed that although flow dipoles in general have sound power proportional to v^6, when they are located well inside a duct the radiated sound power is proportional to v^4 instead. This was borne out by a later study of Nelson and Morfey (1981), who were also able to show that the relationship was different above and below the cut-on frequency of the duct. They were able to collapse their source characteristics at a range of flow rates according to theory for various relative spoiler areas. These collapsed curves did not show the broad peak typical of free jet spectra; instead they fell steadily in amplitude with increasing frequency.

2.2 Mechanical model experiments

One of the earliest efforts to consider fricative production in terms of the aerodynamics involved was that of Meyer-Eppler in 1953. He compared sound-pressure versus flow relationships for plastic tubes with different-sized elliptical openings. He derived effective-width formulas that allowed sound-pressure versus Re plots for different ellipse sizes to coincide, and then used these derived formulas on data of human subjects uttering fricatives to infer articulatory parameters from pressure measurements.

It is not clear, however, that Re_{crit}, which Meyer-Eppler defined to be the lowest Reynolds number at which measurable sound was generated, should be the same for the plastic tubes and the three fricatives /s, ʃ, f/ that he studied. First, Re_{crit} is probably lower for an irregularly shaped vocal tract than for the plastic tubing. Second, sound is most likely being generated for these three fricatives both when the air passes through a constriction (over the tongue for /s, ʃ/ or between teeth and lower lip for /f/) and when the jet of air strikes an obstacle (the teeth for /s, ʃ/ or upper lip for /f/). The intensity of the sound generated may therefore be related more to the distance between the constriction and the obstacle and the physical properties of the obstacle than to the effective area of the constriction. Since the tubes he used had a constriction but no obstacle, application of the effective-width formulas developed for the tubes to the strident fricatives may give misleading results.

Heinz (1958) carried out experiments using a 17 cm tube embedded in a wooden sphere to approximate the dimensions and radiation impedance of a vocal tract. By placing a cylindrical plug with a 0.2 cm diameter axial hole at the mouth and 4 cm back from the mouth, he approximated the articulatory configurations for fricatives such as /ɸ/ and /ç/, respectively. He obtained far-field directivity patterns and spectra for a variety of frequencies and flow rates. As expected, intensity rose with flow rate, except at the half-wavelength resonance of the constriction. Heinz ascribed this behaviour at this resonance to effective movement of the source relative to the constriction. However, the source he used in calculating the system response was a localised pressure source that did not change position with flow rate. The source spectra derived by subtracting the system response from the measured sound spectrum were fairly flat, but with dips at resonance frequencies, which Heinz assumed were a consequence of greater than expected resonance bandwidths due to turbulence losses. He also calculated the incremental flow resistance from his data.

2.3 Theoretical models

Fant (1960) used a distributed model of the vocal tract to investigate the effects of source location, source spectrum, and constriction resistance. Concluding that the theory of turbulence was too undeveloped to be useful, he judged the accuracy of his models by how closely the predicted spectra matched the spectra measured from the speech of a single subject. Models for all fricatives used a series pressure source that generated either white noise for /s, ʃ, ʧ/ or integrated white noise (i.e. with a −6 dB/octave slope) for /x, f/.

Fant did not attempt to make a physical argument relating the two types of source spectra to the distinguishing features of the fricatives. The location of the source – whether at tongue or teeth – is more clearly linked to the place of articulation. The model for /x/, which used a source located at the tongue, produced the best match of all of the fricatives. For /s/ and /ʃ/, although sources at both tongue and teeth were used, it appeared that neither location by itself would provide a good match at all frequencies. Fant suggested that quite possibly /s/ was produced with sources at both locations, with a low spectral level below 1 kHz, but did not attempt a physical justification for this particular source characteristic. Fant was aware that changes in the source location would alter the frequencies of the zeros it produced in the output, but he wrote that this effect would probably prove to be perceptually unimportant.

Flanagan and co-workers (Flanagan, Ishizaka, and Shipley, 1975, 1980; Flanagan and Ishizaka, 1976) elaborated on this model by allowing for multiple noise sources, one for each section of the uniform tube model, with the strength of each source (the variance of the white noise pressure source) depending on Re of the section (computed from its area

and the glottal volume velocity). A given pressure source was included only if its Reynolds number exceeded the value of Re_{crit} determined from Meyer-Eppler (1953). The model allows for a dependence of both intensity and acoustic resistance on flow rate. The sound sources are now distributed throughout the vocal tract, but the source due to a single section is still localised. Further, the spectral characteristic of the source is unchanged by flow rate (except for its overall amplitude), which contradicts findings of Heinz (1958) and Thwaites and Fletcher (1982), among others. Likewise, Re_{crit} is constant regardless of upstream conditions of the tract configuration. The method of computing Re based on the cross-sectional area is not sensitive to the turbulence generated when a jet of air impinges on an obstacle. This model, like Fant's, assumes linear elements, independence of source and filter functions, and one-dimensional sound propagation.

2.4 Speech analysis

Efforts to model fricatives, such as that of Fant (1960), and work with mechanical models, such as that of Heinz (1958), form two approaches to understanding the acoustic mechanisms involved in fricative production. A third approach is simply through acoustic analysis of spoken fricatives. The work of Hughes and Halle (1956), Strevens (1960), Heinz and Stevens (1961) on English, and others – detailed thoroughly by Ladefoged and Maddieson (1986) – is useful but perhaps the most important conclusion is that the considerable variation across speakers makes it difficult to draw any useful generalisations about the acoustic parameters used to differentiate the fricatives.

Some authors pair articulatory and acoustic data in an effort to overcome such problems. Ladefoged and Maddieson exploit not only X-rays, but electropalatography (EPG) and other methods whenever possible. As a result, they are able to make fine articulatory distinctions. However, their descriptions of the spectra still rely on the same crude measures of minimum frequency of the high-energy region, steepness of the slope at cut-on, and the overall amplitude. We will discuss their conclusions regarding the acoustic mechanism for each fricative at a later point. For now, we will say only that by pointing out the full range of articulatory variation possible, they are guaranteeing that there will be a full range of acoustic variation, which we are unlikely to be able to explain at present.

Others are able to proceed further by narrowing the problem considerably. Cohen and Perkell (1986), for instance, used EPG data to demonstrate the effect of the sublingual cavity in distinguishing /s, ʃ/. Hardcastle (1984) has used similar data to study pathological fricatives, and to look at the effect of groove shape on /s, ʃ/.

The approach taken by the author, in light of the above work, has been to focus primarily on experimenting with mechanical models, to relate these occasionally to analysis of human speech, and by this means to test the assumptions on which theoretical models are based (e.g. that source and filter are independent) and determine needed parameter values (e.g. source characteristics). This work has proceeded by means of three sets of experiments with progressively more realistic mechanical models. We will now discuss these models, beginning at the lowest level.

3 Level I Models

3.1 Method

The first set of models was deliberately designed to include only the features thought to be important in fricative production. Thus, the constriction was included, but without significant taper; the vocal tract bend was eliminated, and the transverse cross-section was circular throughout. The models, shown in Figure 3, consisted of a cylindrical plug forming an axial constriction in a tube of circular cross-section; optionally, a semicircular

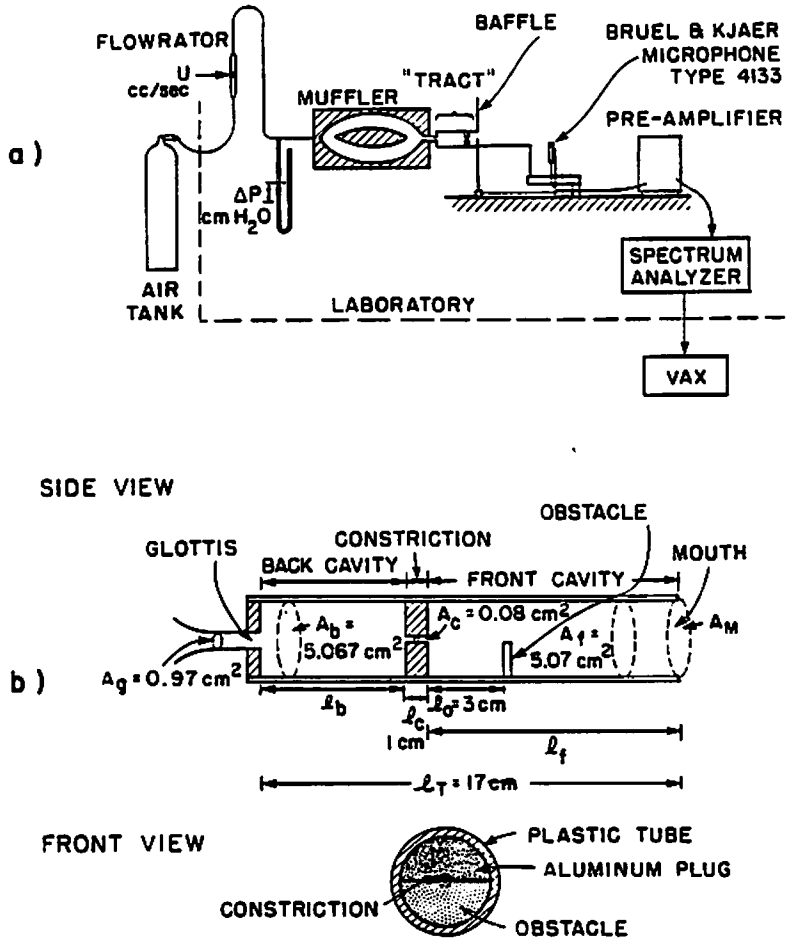

Fig. 3: Cross-sectional diagram of Level I models, with and without obstacle.

obstacle was positioned transverse to the flow downstream of the constriction. Preliminary analysis showed that the articulatory parameters having the greatest acoustic effect were the flow rate, the length of the front cavity, and the presence (or absence) of the obstacle. Acoustic measurements of sound generated by air passing through the mechanical models showed that the obstacle generated relatively high-amplitude noise, which was localised at the obstacle. This led to a division of the models into the *obstacle* and *no-obstacle* cases. Both cases were then examined closely at a range of flow rates for three different front-cavity lengths. Further details of these experiments are given in Shadle (1985, 1986a).

3.2 Results

Theoretical modelling of the obstacle case posited a dipole source at the surface of the obstacle, which has the transmission-line equivalent of a series pressure source. The source characteristic was derived from the experimental data, as shown in Figure 4, and had a rather flat spectrum falling off with frequency; this shape was repeated at higher amplitude for higher flow rate, similar to spoiler-in-duct spectra generated by Nelson and Morfey (1981). When the derived source was used to excite a transmission line whose parameters were determined from the area function of the model according to classic theory, the predicted far-field spectrum matched the experimentally observed spectrum well. This demonstrated that the obstacle case was well-modelled by a series pressure source located at the obstacle, and that that source is independent of the filter. In other words, the source spectrum depends only on flow rate and distance from constriction to obstacle, not on location of the constriction-obstacle unit within the tract.

The no-obstacle case was not modelled theoretically, but it was established that the source should be spatially distributed. Also, if the source was assumed to be determined solely by the flow rate and the dimensions of the constriction, that source was dependent on the location of the constriction within the tract. Thus, altogether, the no-obstacle case is more problematic than the obstacle case.

3.3 Comparison with speech

Comparison of these two models to actual sustained fricatives, as shown in Figure 5, demonstrated that the obstacle case was equivalent to the fricatives /s, ʃ/, with the obstacle playing the role of the teeth. Extensive study of the transfer function of the mechanical model (possible when the articulatory configuration is both simple and completely known) revealed several facts about the mechanical models that could then be extended to the vocal tract for these fricatives. A pressure source located within the front cavity will excite all poles of the entire tract, and in addition a real zero near 0 Hz and a complex-conjugate pair of zeros at a frequency related to the distance between the pressure source and the anterior end of the constriction. Since the distance is so short, a change in it of the order of a millimetre will change the frequency of the free zero by a few hundred hertz. In addition, the pressure source generates zeros that nearly cancel all of the back cavity poles. Thus, the main features remaining in the transfer function are the front cavity poles, the free zero, and the zero near 0 Hz. Because of the low-frequency zero, and the high damping due to the relatively short front cavity, the first formant may be quite low in amplitude.

The no-obstacle case was possibly equivalent to the weak fricatives /ɸ, f, θ/, but due to the high degree of inter-speaker variation, it was difficult to be certain about this claim. Certainly, Catford's data (1977) showing a major drop in spectral energy of /s, ʃ/ but not /θ/ following removal of false teeth support the classification of /s, ʃ/ as obstacle case, and /θ/ as *not* obstacle case. But, as we shall see, the no-obstacle Level I model is not the only alternative to the obstacle case. Thus, the classic theory was shown to be correct if modified in this manner: for /s, ʃ/, a series pressure source located *at the teeth*, with spectral characteristic flat and falling, and having higher amplitude for higher flow rate, is valid. For other fricatives, neither case matched adequately, indicating that the models were perhaps over-idealised.

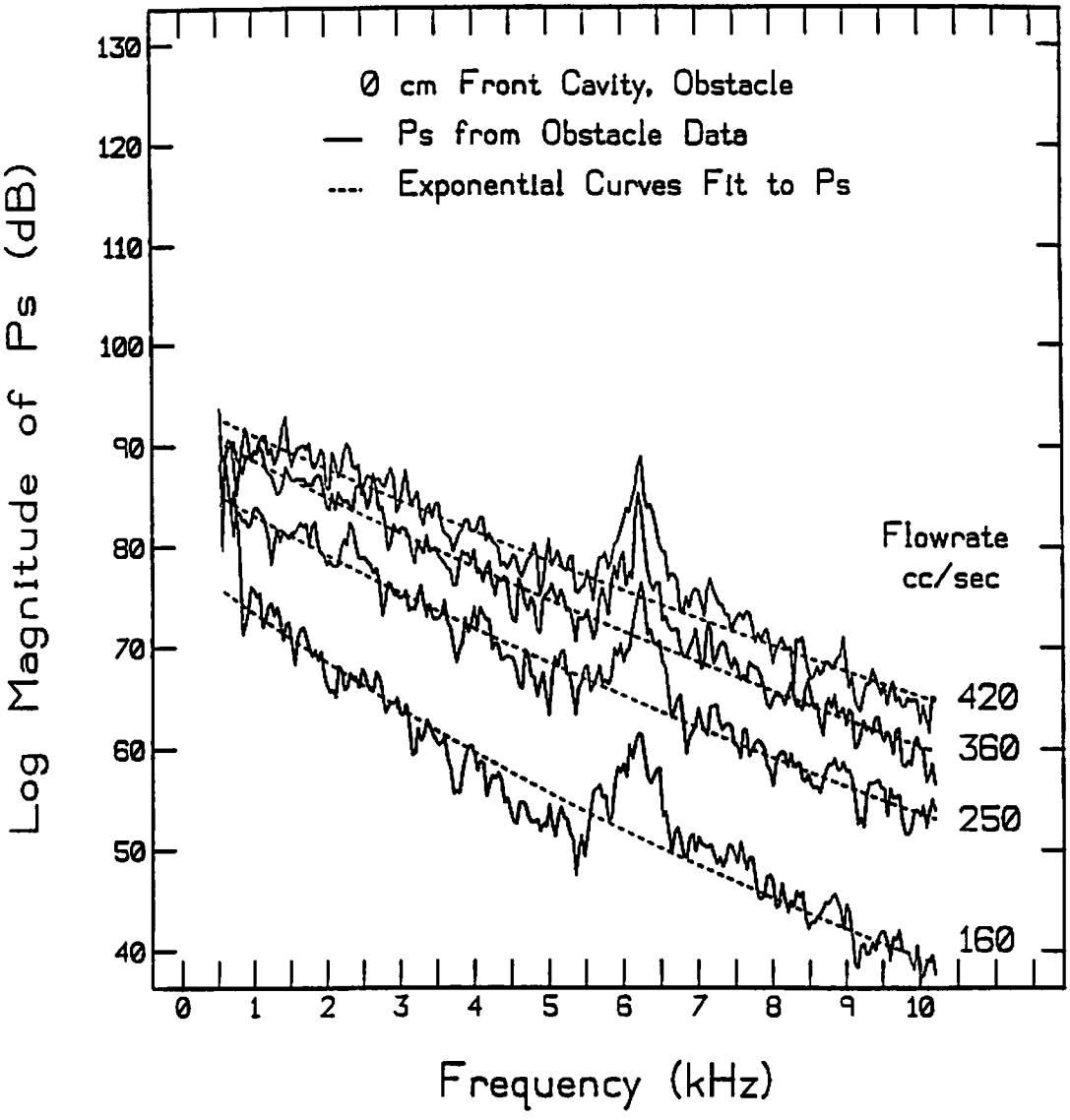

Fig. 4: Equivalent pressure source, p_s, derived from far-field sound pressure. Dashed lines represent an exponential line-fit. Length of front cavity is 0 cm; obstacle is 3 cm downstream of constriction.

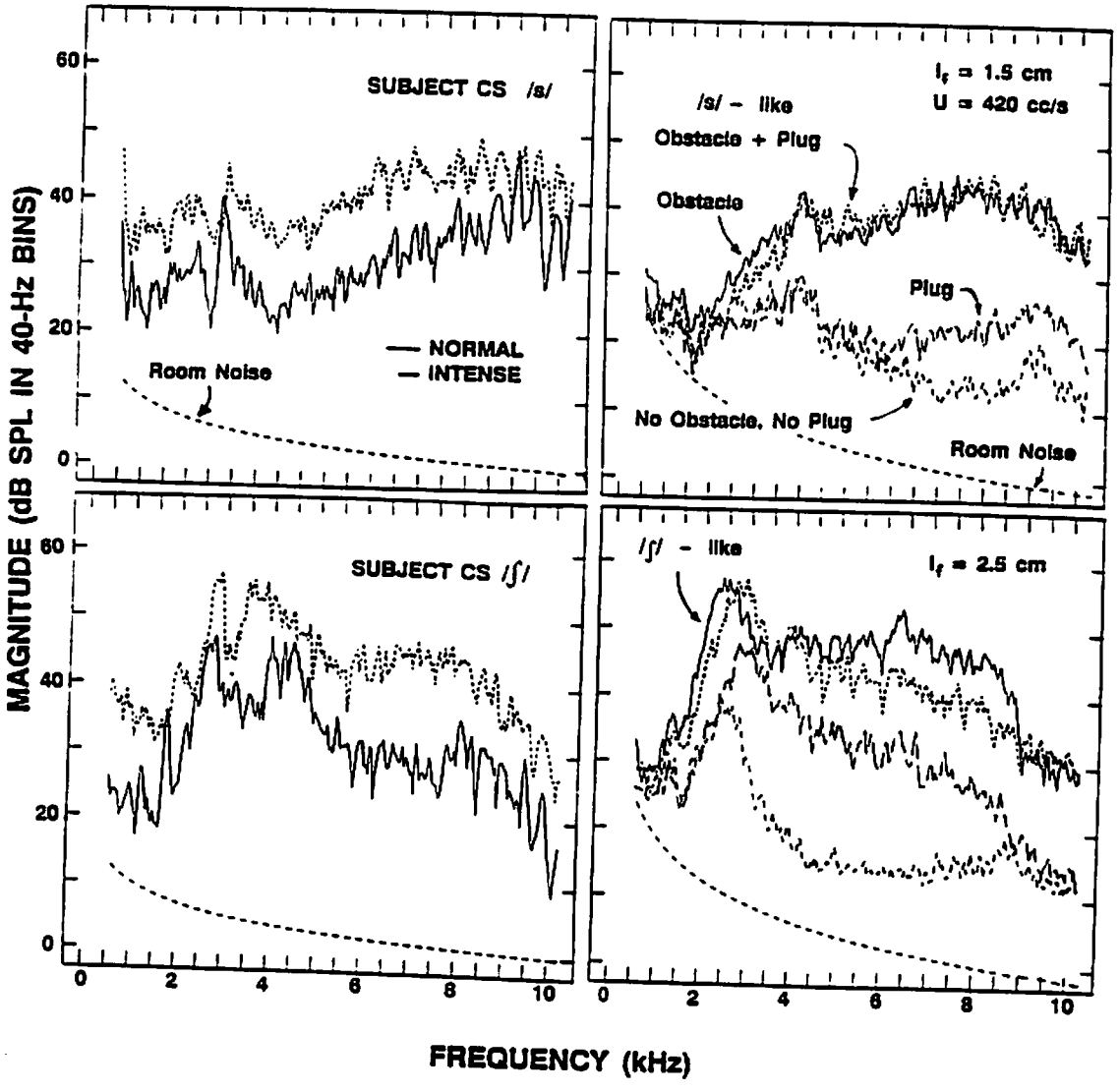

Fig. 5: On left: spectra of /s, ʃ/ by female speaker. On right: spectra generated by the air flowing through obstacle-case models of Fig. 3.

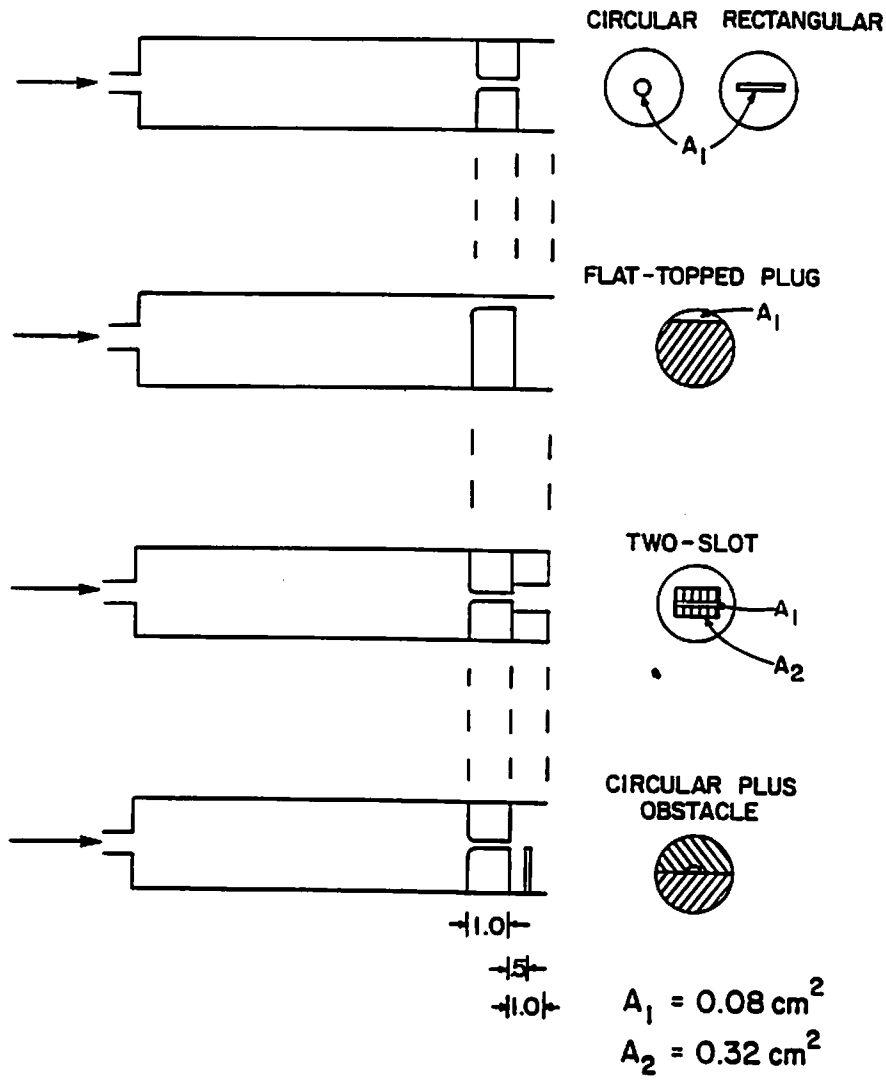

Fig. 6: Cross-sectional and front view of Level II models.

4 Level II Models

4.1 Method

Fricatives other than /s, ʃ/ were investigated using Level II models, which featured constriction shapes that were non-circular in the transverse plane, as shown in Figure 6. These were used in the Level I setup shown in Figure 3, and as before, spectra were computed at a variety of flow rates and constriction locations. Full details are given in Shadle (1986b).

4.2 Results

Comparison of speech spectra and model output showed that /ç, x/ were best matched by a 'flat-topped' constriction which forced the air to flow along the wall of the tube, as shown in Figure 7. This generated more sound than a centred circular constriction of the same cross-sectional area and length, and showed even greater dependence on the location of the constriction within the tube. Thus, the demonstrated source-tract dependence of the no-obstacle case appeared to be due to sound generation along the walls of the tube downstream of the constriction. However, theoretical models of the flat-topped plug indicated that the similarity to speech might possibly be due to an artifact (Shadle, 1986b). Essentially, Level II models were so highly idealised that it was unclear whether the same physical mechanisms for sound production were involved.

The Level II models did, however, allow clarification of the problem as follows. Since the self-noise of turbulent air is much lower in amplitude than that generated by air impinging on a rigid object, what objects does the air impinge upon in its path through the vocal tract? How do these objects affect the sound generated?

5 Level III Models

5.1 Method

To answer these questions, more realistic models of the vocal tract were needed. These Level III fricative models were designed and built in the following manner. The mid-sagittal X-ray tracings of fricatives given by Fant (1970) were used as a pattern with which to cut 'tracts' out of inch-thick (2.54 cm) plexiglass. The 'tract' was then sandwiched between two layers of $\frac{1}{4}$ inch thick plexiglass, and the entire sandwich bolted together. The overall area function was thus on average correct, but at specific locations along the tract, incorrect, since there was a constant 1 inch width throughout. This was deemed unworkable only in the region of the constriction, where the area was corrected by the insertion of modelling clay to create a central channel of length 0.5 cm and width 0.1 cm. The variations in height of the models at the point of greatest constriction led to cross-sectional areas of between 0.015 and 0.03 cm^2, which is actually smaller than typical fricative constrictions. Effectively this would tend to generate more sound at a given flow rate; since a range of flow rates was investigated, the difference was not deemed problematic.

In the completed model, holes were then drilled in the vocal tract walls at various points downstream of the constriction to allow a Bruel & Kjaer 4170 probe microphone to be inserted and lie flush with the inner wall of the tract. The tip of the probe microphone was 1.25 mm in diameter. The tapered construction of the probe provides an impedance-matching function so that there are no resonances of the probe itself interfering with the acoustic signal. Since only one probe microphone was used, aluminium plugs were machined to fit the probe holes, and were inserted when those holes were not in use.

The fricative models thus created corresponded to the fricatives /ʃ, ç, x/, and are shown in Figure 8. /ç/ was actually created from an X-ray tracing of the glide /j/, which has the same place of articulation. Note that on all fricatives the uvulas have been removed, since this is a feature that does not extend uniformly across the tract.

These fricative models were then placed, one at a time, in an anechoic chamber. Figure 9 shows the experimental setup. Air passed through the model at a range of flow rates (330 to 1670 cm^2/s). The air came from an air compressor, through a pressure-regulating valve, through a rotameter to measure the flow rate, and then into the anechoic chamber. There, the 4 cm air duct entered a large L-shaped muffler constructed of perforated piping inside a foam-lined plywood box. The purpose of the muffler was to remove the sound

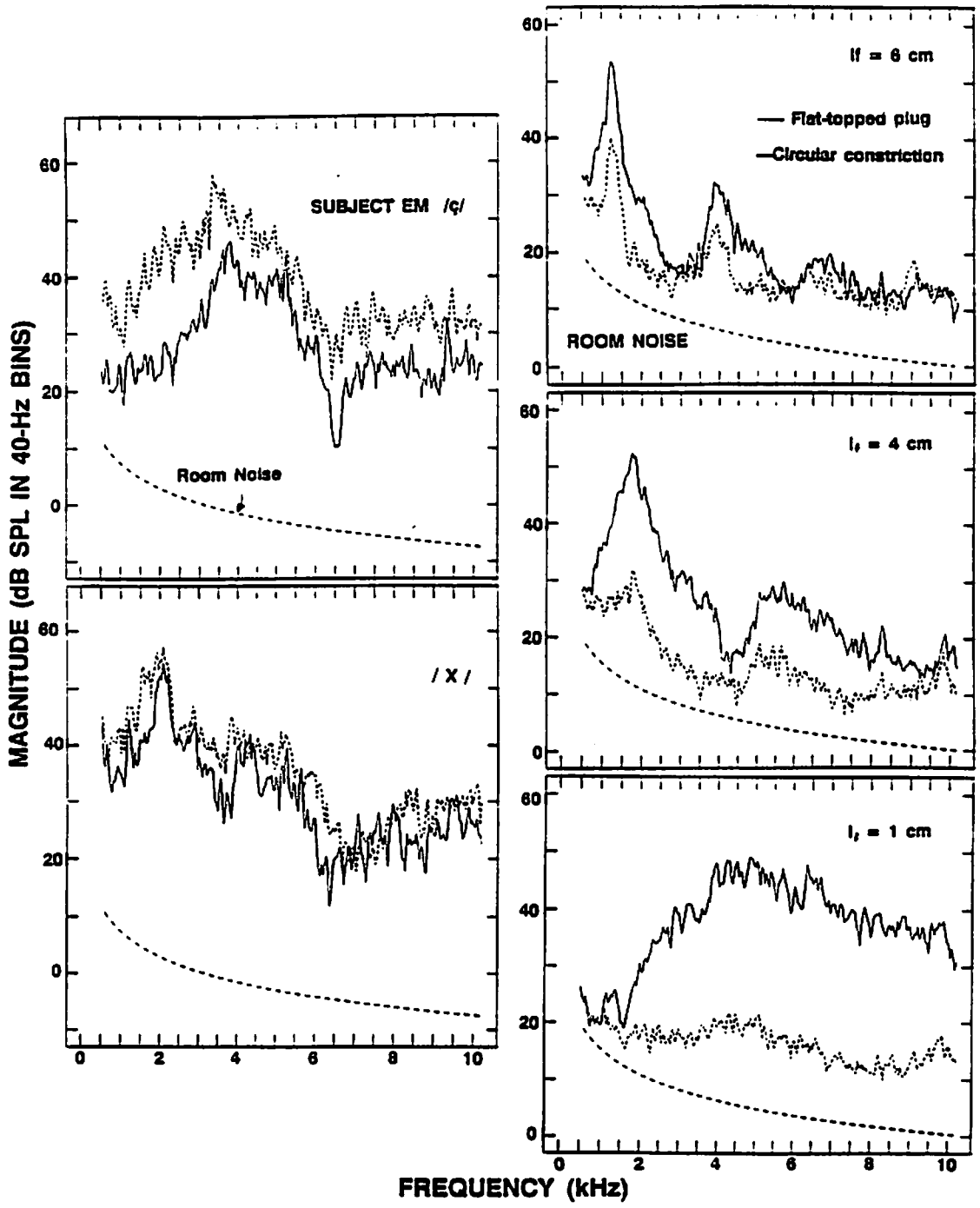

Fig. 7: On left: spectra of /ç, x/ by female speaker. Solid lines are for normal production and dotted lines for intense production. On right: spectra produced by the indicated constrictions and lengths of front cavity, at flow rate 400 cm^3/s.

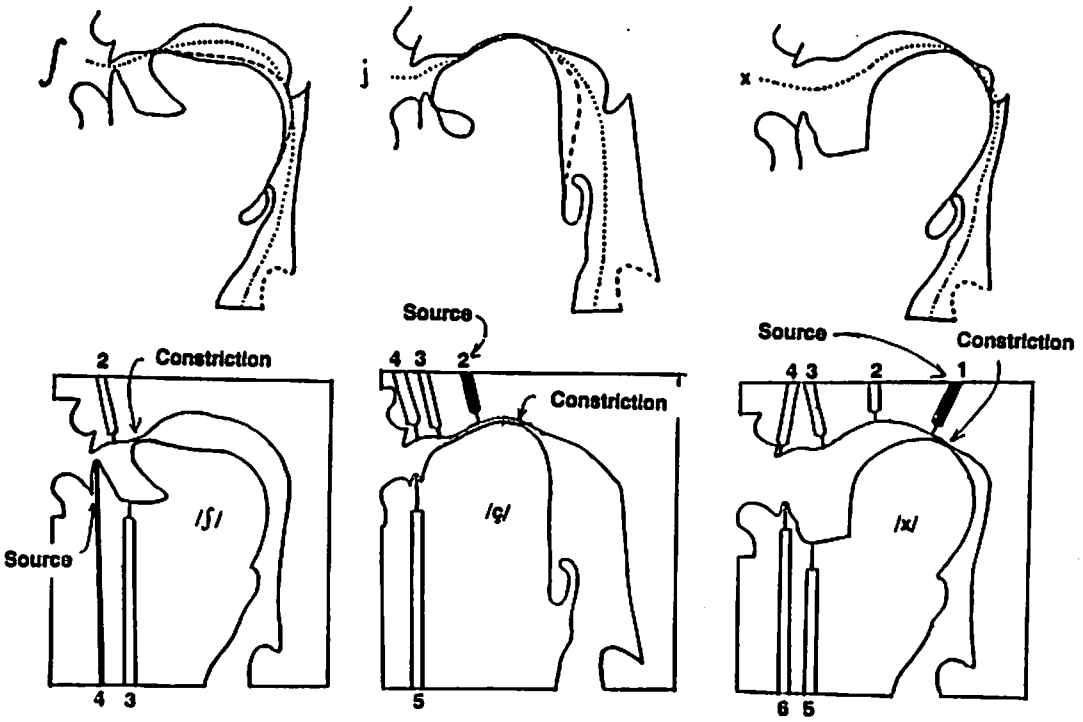

Fig. 8: Level III models, showing probe holes, and the X-rays from which they were derived. The active source points are labelled. Arrows marked "Constriction" point to region of smallest cross-sectional area in the tract.

generated in the upstream plumbing and to reduce upstream reflections of the sound from the model. The fricative model was attached directly to the output of the muffler, and was thus positioned in roughly the centre of the anechoic chamber.

The sound pressure at the probe microphone, and at a far-field $\frac{1}{2}$ inch B&K 4133 condenser microphone, was analysed by a B&K 2032 spectrum analyser. For each combination of fricative, probe location, and flow rate, three functions were computed by the spectrum analyser and graphed. The averaged power spectrum was computed for each microphone, averaging 100 spectra with no overlap. A 62.5 ms Hanning window was used at a 25.6 kHz sampling rate. The preamplifier characteristics were factored in so that the graphs shown represent absolute sound pressure level. In addition, the coherence was computed from the two averaged power spectra.

5.2 Results

5.2 a Source location

An example of the power spectra and coherence for two probe locations is shown in Figure 10. When the two microphone signals were highly coherent (i.e. close to 1.0), sound at both microphones was similar, which indicates that in this case (probe point 3) the probe microphone was detecting the tract resonances only. When the signals were incoherent (i.e. close to 0.0), the sound at both microphones was dissimilar, indicating that the probe microphone was detecting noise where it was generated. Thus, in this case

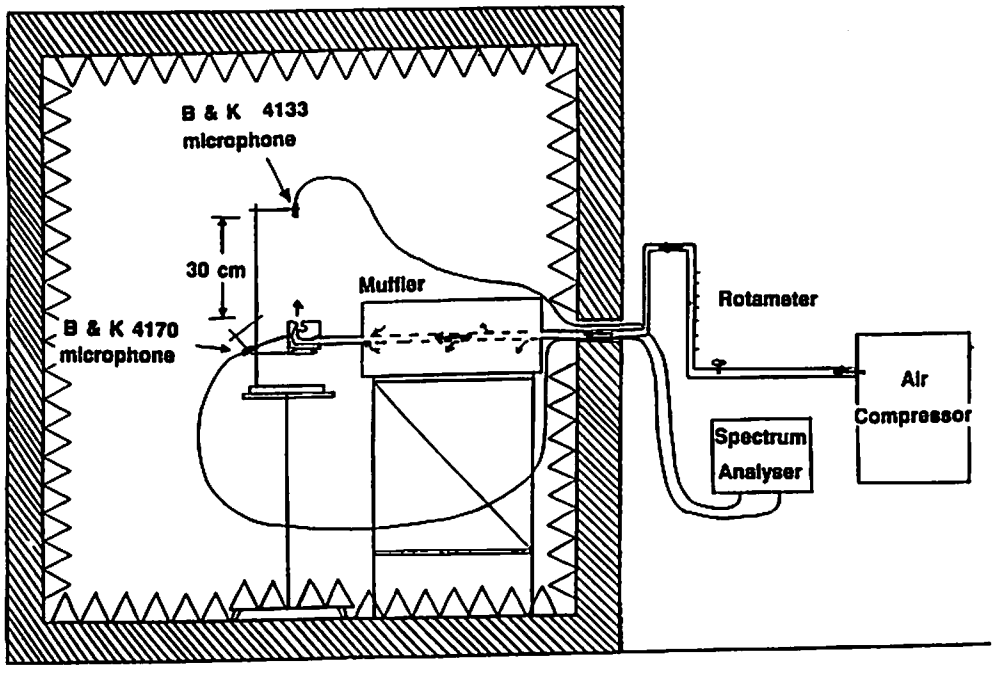

Fig. 9: Diagram of the experimental setup for Level III models.

(probe point 4) the probe was at or near a source location. These comparisons were made for every fricative, every probe point, and at four different flow rates (330, 670, 1000 and 1330 cm^2/s).

In Figure 8, the probe locations labelled "source" are those for which the microphone signals were generally incoherent across the frequency range. The same locations proved to be 'active' source locations across the range of flow rates used. Note the differences between the fricatives: for /ʃ/ the active probe point is at the lower teeth, which Level I model results had led us to expect to act as an obstacle; for /ç, x/ the active probe point is in each case the point nearest to the constriction, located along the wall rather than on any downstream projection.

Due to the construction of the models, only a small number of probe points could be investigated. It seems likely that if the probe position could be varied continually, the results for /ʃ/ would stay substantially the same, but /ç, x/ would prove to have an active source *region* rather than point. We will consider this again in Section 5.3.

5.2 b Source characteristic

At each active source point, the spectrum was analyzed at nine different flow rates (from 330 to 1670 cm^2/s). Shown in Figure 11 are the families of curves for each of the fricatives /ʃ, ç, x/.

Differences can be classified according to three spectral attributes. The first attribute, source spectrum shape, is flat although decreasing in amplitude at increased frequency for /ʃ/, but possesses a broad peak for /ç, x/. The peak frequency ranges from 4 to 7 kHz for /ç/, or 5 to 9 kHz for /x/; as flow rate increases, the peak frequency increases. This

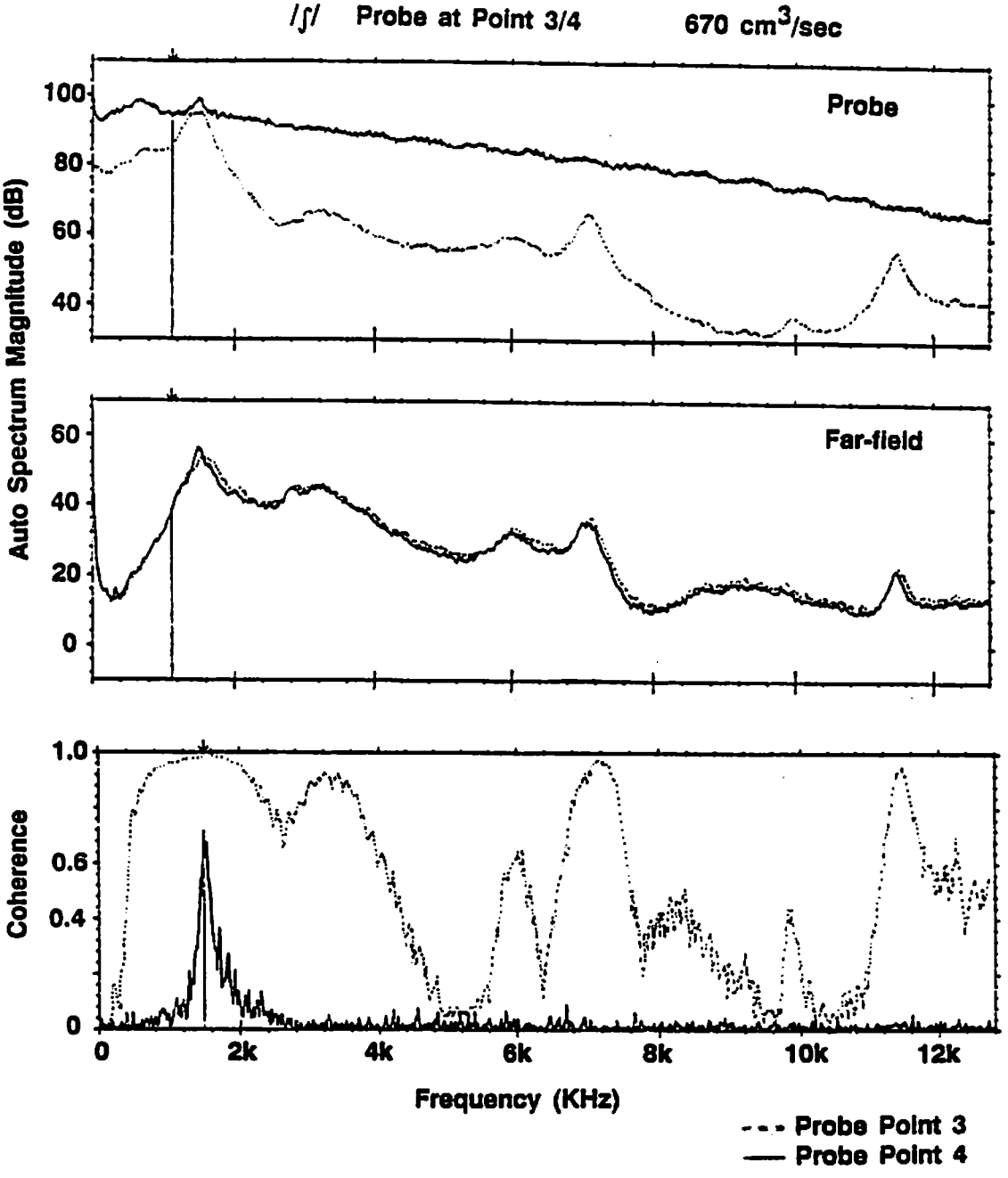

Fig. 10: Example of the spectrum analyzer output for /ʃ/, for probe point 3 (floor of mouth, just posterior of lower teeth) and point 4 (tip of lower teeth).

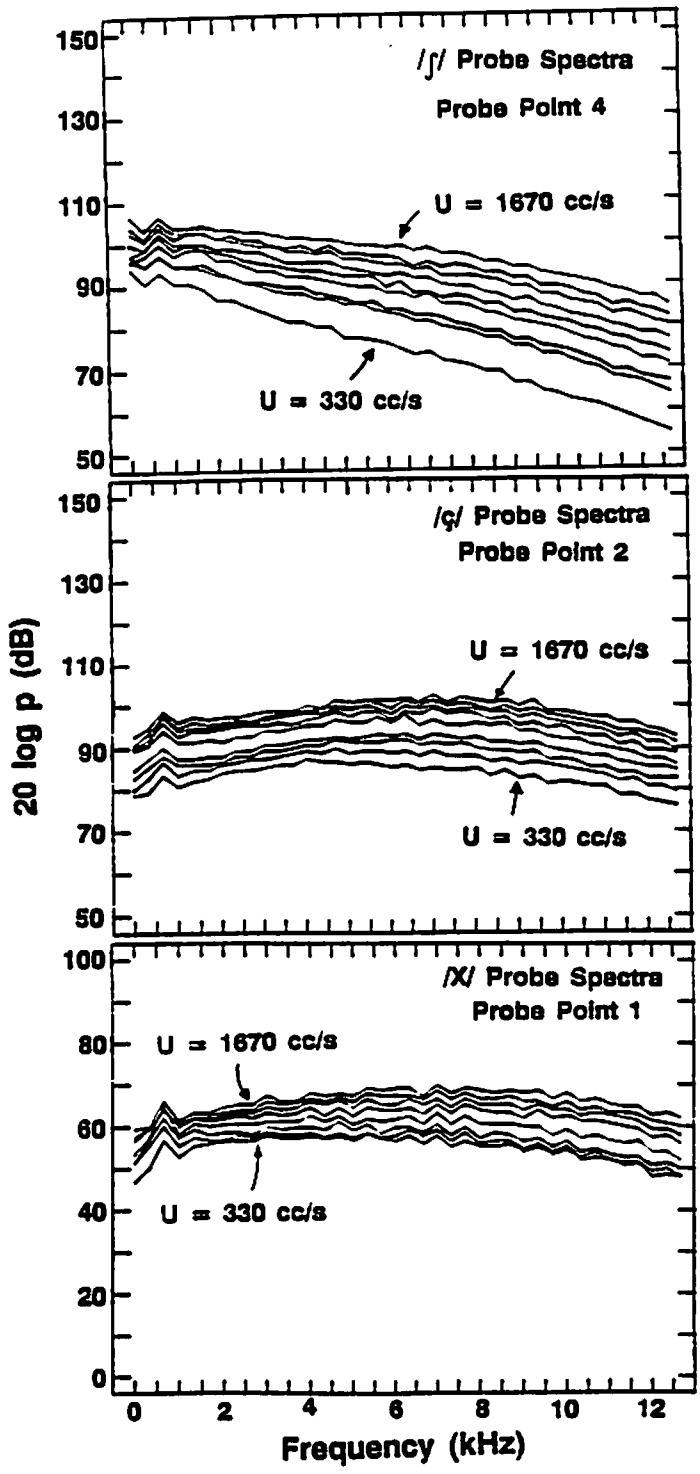

Fig. 11: Families of source spectra for each of the three fricatives. These are for the output of the probe microphone while positioned in the active source point, for the indicated range of flow rates.

peak-frequency increase is predicted from results for noise of free jets, although the jets emerging from the constrictions in these fricative models cannot be considered to be free. The flat falling spectrum for /ʃ/ is similar to those derived for the Level I obstacle-case models, and to those derived by Nelson and Morfey.

The second spectral attribute is source spectrum amplitude. Since the spectra preserve absolute sound pressure levels, it is clear that the /ʃ/ source has the highest amplitude, /x/ the lowest, at any given flow rate. These are inversely related to distance of the source location from constriction, although the constriction area is not identical in all cases, confounding results. Therefore it is not known at this point if that relation will hold across further configurations. It does seem likely that the obstacle source, as for /ʃ/, is intrinsically more efficient and therefore of higher amplitude, in this flow rate range, than the 'wall' source.

The third spectral attribute is the rate of change of amplitude with flow rate. This is not uniform across the entire frequency range, nor is it expected to be. If we consider the difference in amplitude from lowest to highest flow rate at 2 and at 12 kHz, we obtain a rough measure of the dependence of amplitude on flow rate. For /ʃ/, the difference is 17 dB at 2 kHz, 29 dB at 12 kHz; for /ç/, 13 and 18 dB; for /x/, 8 and 15 dB. Since the rate of change of amplitude is one of the prime characteristics differentiating flow noise source types, these differences are significant, and point once more to a different source mechanism at work.

In summary, the properties of source location and spectral shape group /ç, x/ together, distinguishing them from /ʃ/; spectral amplitude and rate of change of amplitude with flow rate separate the fricatives, with /ʃ/ having the maximum values, and /x/ the minimum. Thus, the search for the source properties has revealed distinguishing acoustic features.

5.3 Comparison with speech

Because these models were based quite closely on an actual vocal tract, it was possible to make detailed comparisons of the model output with the human speech analyzed by Fant. Figure 12 shows the results: in all three cases, the output spectra were significantly dissimilar. Because of the constant width throughout the mechanical model, however, the human and mechanical area functions are not the same. VCTR, a computer program that calculates the transfer function given an area function and various loss factors (Badin and Fant, 1984), was used to explore whether the area function differences would explain the acoustic differences.

The predicted transfer functions are shown in the bottom graph of each section of Figure 12. Adjusting the area function, and the constriction-source distance with it, did explain the discrepancies for /ʃ/. However, the best explanation for the discrepancies in /ç, x/ appeared to be that the source is distributed. This would blur the free zero and, in the case of /x/, alter the relative amplitude of the first and second formants. Although this could be tested directly, we now have stronger evidence that the wall source is distributed spatially, sufficiently to affect the acoustic output.

At this point it is worthwhile to return to Ladefoged and Maddieson (1986), who identified every fricative as belonging to the obstacle or no-obstacle case, which they declared equivalent to sibilant and non-sibilant. It may seem a matter of semantics only to refer to a 'wall' case instead of a 'no-obstacle' case. This is not so, however. The turbulent jet of air generates additional sound at any rigid boundaries it encounters. The difference between a wall and an obstacle is a matter only of angle with respect to the jet axis, and could theoretically fall along a continuum. Further, sound generation will not necessarily take place at all rigid boundaries occurring at right angles to the flow;

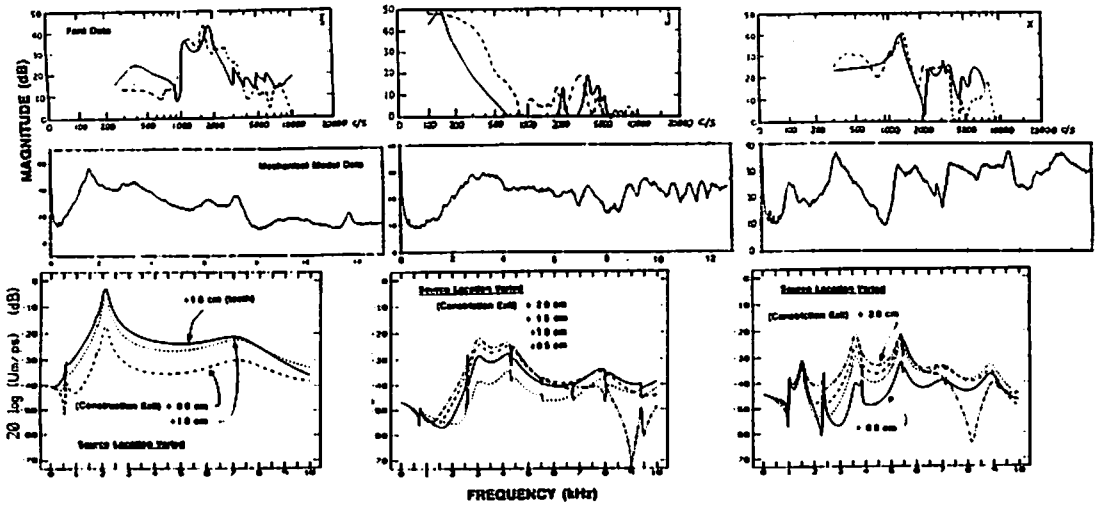

Fig. 12: For each fricative shown, the acoustic output measured by Fant (1960), that produced by the mechanical model with the mid-sagittal profile of Fant's subject, and transfer functions derived using VCTR on the area functions are indicated. Note that Fant uses the symbol š where we have used ʃ.

that is, not all objects will act as obstacles. The parameters of velocity, angle and area of the jet at the surface will determine the source type and parameters. It thus seems a bit premature to label each fricative sibilant or non-sibilant according to the position of the lower teeth during its production, although that will certainly eventually be an important parameter defining the source.

6 Conclusion

Based on the work reported above, it appears that sound generation in fricative consonants occurs in at least two different ways. We refer to these as the obstacle source and the wall source, terms derived from the distinguishing articulatory feature.

The obstacle source occurs in cases where principal sound generation is at a rigid body that is approximately normal to the flow. For /ʃ/, the obstacle appears to be the lower teeth; /s/ is also an example of an obstacle source, with the obstacle possibly at upper rather than lower teeth. The obstacle source is characterised by maximum source amplitude for a given flow velocity, by a spectral shape that is flat and falling off with increased frequency, and by a maximum rate of change of sound pressure with volume velocity. It can be modelled with a series pressure source at the obstacle. It appears that the air must be travelling above a certain critical velocity when it impinges on the obstacle, implying a certain relationship among the parameters constriction area, flow rate (or pressure drop across the constriction), and distance to the obstacle from the constriction. The details of this relationship are not currently known.

The wall source occurs in cases where the principal sound generation is along a relatively rigid wall, approximately parallel to the flow. The fricatives /ç, x/ use a wall source, and it appears likely that many other fricatives do as well. It is characterised by high, but

not maximum, source amplitude for a given flow rate, by a spectral shape with a broad peak, and by a high, but not maximum, rate of change of sound pressure with volume velocity. It appears likely that it is in fact a distributed source; details of the coherence of different parts of the source are unknown at present. The distributed source should be located at the point in the model corresponding to just downstream of the constriction.

It appears that by considering the three-dimensional shape of the tract in order to characterise the source, we can obtain a one-dimensional model in which the source is independent of the tract. This allows us to use superposition and linear system theory, which is a substantial simplification. However, this may require a fairly complex model of the source.

7 Acknowledgements

This work was supported by a Hunt Fellowship awarded by the Acoustical Society of America, and a grant awarded by the North Atlantic Treaty Organization, in 1984.

The material in this Report is to appear in the Proceedings of the NATO Study Institute on Speech Production, held at Bonas, France, 16–20 July 1989, which are to be published by Kluwer Academic of Dordrecht.

8 References

BADIN, P. AND FANT, G. (1984) "Notes on vocal tract computation", *Speech Transmission Laboratory Quarterly Status and Progress Report, STL-QPSR 2-3/1984, KTH, Stockholm*, 53–108.

CATFORD, J.C. (1977) *Fundamental Problems in Phonetics*, Indiana University Press, Bloomington, IN.

COHEN, M. AND PERKELL, J. (1986) "Palatographic and acoustic measurements of the fricative consonant pair /s/ and /ʃ/", *Proceedings of the 12th International Congress on Acoustics*, Toronto, Paper A3–5.

FANT, C.G.M. (1960) *The Acoustic Theory of Speech Production*, Mouton, The Hague.

FLANAGAN, J.L. AND ISHIZAKA, K. (1976) "Automatic generation of voiceless excitation in a vocal cord-vocal tract speech synthesizer", *IEEE Transactions on Acoustics, Speech and Signal Processing*, **ASSP-24**, 163–170.

FLANAGAN, J.L., ISHIZAKA, K. AND SHIPLEY, K.L. (1980) "Signal models for low bit-rate encoding of speech", *Journal of the Acoustical Society of America*, **68**, 780–791.

FLANAGAN, J.L., ISHIZAKA, K. AND SHIPLEY, K.L. (1975) "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", *Bell System Technical Journal*, **54**, 485–506.

FLETCHER, N. AND THWAITES, S. (1983) "The physics of organ pipes", *Scientific American*, **248**, 94–103.

GOLDSTEIN, M.E. (1976) *Aeroacoustics*, McGraw-Hill, New York, NY.

HARDCASTLE, W.J. (1984) "New methods of profiling lingual palatal contact patterns with electropalatography", *University of Reading, Work in Progress, No. 4*, 1–40.

HEINZ, J.M. (1958) "Sound generation by turbulent flow in an acoustic resonator", *MS Thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA.*

HEINZ, J.M. AND STEVENS, K.N. (1961) "On the properties of voiceless fricative consonants", *Journal of the Acoustical Society of America*, **33**, 589–596.

HELLER, H.H. AND WIDNALL, S.E. (1970) "Sound radiation from rigid flow spoilers correlated with fluctuating forces", *Journal of the Acoustical Society of America*, **47**, 924–936.

HUGHES, G. AND HALLE, M. (1956) "Spectral properties of fricative consonants", *Journal of the Acoustical Society of America*, **28**, 303–310.

LADEFOGED, P. AND MADDIESON, I. (1986) "Some of the sounds of the world's languages", *UCLA Working Papers in Phonetics, No. 64.*

LIGHTHILL, M.J. (1952) "On sound generated aerodynamically: I. General theory", *Proceedings of the Royal Society (London)*, **A211**, 564–587.

LIGHTHILL, M.J. (1954) "On sound generated aerodynamically: II. Turbulence as a source of sound", *Proceedings of the Royal Society (London)*, **A222**, 1–32.

MEYER-EPPLER, W. (1953) "Zum Erzeugungsmechanismus der Geräuschlaute", *Zeitschrift für Phonetik*, **7**, 196–212.

MORSE, P.M. AND INGRAD, K.U. (1968) *Theoretical Acoustics*, McGraw-Hill, New York, NY.

NELSON, P.A. AND MORFEY, C.L. (1981) "Aerodynamic sound production in low speed flow ducts", *Journal of Sound and Vibration*, **79**, 263–289.

SCHLICHTING, H. (1979) *Boundary Layer Theory (7th edition)*, McGraw-Hill, New York, NY.

SHADLE, C.H. (1985) "The acoustics of fricative consonants", *PhD Thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA*; also released as *Research Laboratory of Electronics Technical Report 506.*

SHADLE, C.H. (1986a) "Models of turbulent noise sources in the vocal tract", *Proceedings of the Institute of Acoustics*, **8**:3, 213–220.

SHADLE, C.H. (1986b) "Models of fricative consonants involving sound generation along the wall of a tube", *Proceedings of the International Congress on Acoustics*, Toronto, Paper A3–4.

STREVENS, P. (1960) "Spectra of fricative noise in human speech", *Language and Speech*, **3**, 32–49.

THWAITES, S. AND FLETCHER, N.H. (1982) "Wave propagation on turbulent jets: II. Growth", *Acustica*, **51**, 44–49.